WDL-RF: Predicting Bioactivities of Ligand Molecules Acting with G Protein-coupled Receptors by Combining Weighted Deep Learning and Random Forest

Jiansheng Wu, Qiuming Zhang, Weijian Wu, Tao Pang, Haifeng Hu, Wallace K.B. Chan, Xiaoyan Ke, Yang Zhang

Supporting Information

Table S1. List of classes and subfamilies of GPCR datasets used in this study.

<table>
<thead>
<tr>
<th>Class</th>
<th>Subfamilies</th>
<th>GPCR ID (# of Known Ligands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Aminergic receptors</td>
<td>P08908(2294), P50406(1421), P08912(369), P35348(1027), P21917(1679), Q9Y5N1(2092),</td>
</tr>
<tr>
<td></td>
<td>Peptide receptors</td>
<td>P30968(1124), P24530(1019), Q99705(2052), P35372(3828), P46663(452), P35346(689), P21452(696),</td>
</tr>
<tr>
<td></td>
<td>Nucleotide receptors</td>
<td>P30542 (3016),</td>
</tr>
<tr>
<td></td>
<td>Lipid receptors</td>
<td>Q95000(317), Q9Y5Y4(641), P34995(236),</td>
</tr>
<tr>
<td></td>
<td>Protein receptors</td>
<td>P51677(781),</td>
</tr>
<tr>
<td></td>
<td>Melatonin receptors</td>
<td>P48039(684),</td>
</tr>
<tr>
<td></td>
<td>Steroid receptors</td>
<td>Q8TDU6(1153),</td>
</tr>
<tr>
<td></td>
<td>Alicarboxylic acid receptors</td>
<td>Q8TDS4(271),</td>
</tr>
<tr>
<td></td>
<td>Sensory receptors</td>
<td>no known ligands for all of these GPCRs</td>
</tr>
<tr>
<td></td>
<td>Orphan receptors</td>
<td>Q9HC97(1589),</td>
</tr>
<tr>
<td>B</td>
<td>Peptide receptors</td>
<td>P47871(1129),</td>
</tr>
<tr>
<td></td>
<td>Adhesion receptors</td>
<td>no known ligands for all of these GPCRs</td>
</tr>
<tr>
<td>C</td>
<td>Ion receptors</td>
<td>P41180(940),</td>
</tr>
<tr>
<td></td>
<td>Amino acid receptors</td>
<td>Q14416(1810),</td>
</tr>
<tr>
<td></td>
<td>Sensory receptors</td>
<td>no known ligands for all of these GPCRs</td>
</tr>
<tr>
<td></td>
<td>Orphan receptors</td>
<td>no known ligands for all of these GPCRs</td>
</tr>
<tr>
<td>F</td>
<td>Protein subfamily</td>
<td>Q99835(1523),</td>
</tr>
<tr>
<td>Taste 2</td>
<td>Sensory receptors</td>
<td>no known ligands for all of these GPCRs</td>
</tr>
<tr>
<td>Other</td>
<td>Orphan receptors</td>
<td>no known ligands for all of these GPCRs</td>
</tr>
</tbody>
</table>
Table S2. Performance of Molecular Fingerprints from different module units.

<table>
<thead>
<tr>
<th>EC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Layer&lt;sup&gt;b&lt;/sup&gt;</th>
<th>GPCR Datasets</th>
<th>P30968</th>
<th>P24530</th>
<th>Q99705</th>
<th>P35372</th>
<th>P30542</th>
<th>P08908</th>
<th>P50406</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WDL</td>
<td></td>
<td>1.21</td>
<td>1.04</td>
<td>1.33</td>
<td>1.51</td>
<td>0.96</td>
<td>1.50</td>
<td>1.45</td>
</tr>
<tr>
<td>RMSE (↓)</td>
<td>1st</td>
<td></td>
<td>2.02*</td>
<td>1.95*</td>
<td>2.22*</td>
<td>2.35*</td>
<td>1.95*</td>
<td>2.31*</td>
<td>2.52*</td>
</tr>
<tr>
<td></td>
<td>2nd</td>
<td></td>
<td>1.99*</td>
<td>1.81*</td>
<td>2.36*</td>
<td>2.21*</td>
<td>1.90*</td>
<td>2.17*</td>
<td>2.57*</td>
</tr>
<tr>
<td></td>
<td>3rd</td>
<td></td>
<td>1.75*</td>
<td>1.79*</td>
<td>2.12*</td>
<td>2.13*</td>
<td>1.81*</td>
<td>2.07*</td>
<td>2.48*</td>
</tr>
<tr>
<td></td>
<td>4th</td>
<td></td>
<td>1.31*</td>
<td>1.24*</td>
<td>1.50*</td>
<td>1.60</td>
<td>0.98</td>
<td>1.55</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td>5th</td>
<td></td>
<td>1.59*</td>
<td>1.73*</td>
<td>2.03*</td>
<td>1.98*</td>
<td>1.69*</td>
<td>1.97*</td>
<td>2.41*</td>
</tr>
<tr>
<td></td>
<td>6th</td>
<td></td>
<td>1.61*</td>
<td>1.71*</td>
<td>1.99*</td>
<td>2.01*</td>
<td>1.69*</td>
<td>2.01*</td>
<td>2.39*</td>
</tr>
<tr>
<td></td>
<td>WDL</td>
<td></td>
<td>0.86</td>
<td>0.85</td>
<td>0.82</td>
<td>0.74</td>
<td>0.86</td>
<td>0.70</td>
<td>0.79</td>
</tr>
<tr>
<td>r&lt;sup&gt;2&lt;/sup&gt;↑</td>
<td>1st</td>
<td></td>
<td>0.58*</td>
<td>0.44*</td>
<td>0.49*</td>
<td>0.36*</td>
<td>0.41*</td>
<td>0.27*</td>
<td>0.38*</td>
</tr>
<tr>
<td></td>
<td>2nd</td>
<td></td>
<td>0.60*</td>
<td>0.52*</td>
<td>0.41*</td>
<td>0.44*</td>
<td>0.44*</td>
<td>0.37*</td>
<td>0.36*</td>
</tr>
<tr>
<td></td>
<td>3rd</td>
<td></td>
<td>0.70*</td>
<td>0.55*</td>
<td>0.54*</td>
<td>0.48*</td>
<td>0.50*</td>
<td>0.47*</td>
<td>0.40*</td>
</tr>
<tr>
<td></td>
<td>4th</td>
<td></td>
<td>0.84</td>
<td>0.77*</td>
<td>0.77</td>
<td>0.71</td>
<td>0.85</td>
<td>0.67</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>5th</td>
<td></td>
<td>0.76*</td>
<td>0.59*</td>
<td>0.57*</td>
<td>0.51*</td>
<td>0.55*</td>
<td>0.53*</td>
<td>0.42*</td>
</tr>
<tr>
<td></td>
<td>6th</td>
<td></td>
<td>0.74*</td>
<td>0.61*</td>
<td>0.59*</td>
<td>0.51*</td>
<td>0.53*</td>
<td>0.51*</td>
<td>0.45*</td>
</tr>
<tr>
<td>q&lt;sup&gt;2&lt;/sup&gt;↑</td>
<td>1st</td>
<td></td>
<td>0.58*</td>
<td>0.43*</td>
<td>0.48*</td>
<td>0.34*</td>
<td>0.40*</td>
<td>0.27*</td>
<td>0.37*</td>
</tr>
<tr>
<td></td>
<td>2nd</td>
<td></td>
<td>0.60*</td>
<td>0.50*</td>
<td>0.40*</td>
<td>0.42*</td>
<td>0.43*</td>
<td>0.34*</td>
<td>0.34*</td>
</tr>
<tr>
<td></td>
<td>3rd</td>
<td></td>
<td>0.69*</td>
<td>0.51*</td>
<td>0.52*</td>
<td>0.46*</td>
<td>0.48*</td>
<td>0.41*</td>
<td>0.39*</td>
</tr>
<tr>
<td></td>
<td>4th</td>
<td></td>
<td>0.82</td>
<td>0.75*</td>
<td>0.76</td>
<td>0.70</td>
<td>0.84</td>
<td>0.67</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>5th</td>
<td></td>
<td>0.75*</td>
<td>0.57*</td>
<td>0.55*</td>
<td>0.50*</td>
<td>0.53*</td>
<td>0.54*</td>
<td>0.41*</td>
</tr>
<tr>
<td></td>
<td>6th</td>
<td></td>
<td>0.72*</td>
<td>0.56*</td>
<td>0.58*</td>
<td>0.49*</td>
<td>0.50*</td>
<td>0.53*</td>
<td>0.42*</td>
</tr>
</tbody>
</table>

<sup>a</sup>Evaluation Criterion: ↑ (↓) indicates the larger (smaller), the better the model performance; the best results on each evaluation criterion are highlighted in boldface.

<sup>b</sup>Layers: The WDL is the default molecular fingerprint with four module units used in this paper, i.e., the weighted molecular fingerprint of all module units. The 1st to 6th respectively denote the molecular fingerprint generated by the first to the sixth module unit of our approach WDL-RF.

GPCR datasets: '*' indicates the performance of the molecular fingerprint from a module unit is significantly worse than that of WDL based on Wilcoxon signed-ranked test.